**Module Name: BUS14373**

**Student ID: 20325050**

## Part I

## Executive Summary

Customer churn is a concept where the customer stops to buy the product or service from the company. There is no consensus on the definition of churn as time for repeat purchase differ throughout different industries. However, it is shorter for retail stores as they sell regular household items. Keeping in mind the nature of Food Corp business, a period of 28 days is set for this study to be the churning period. Consulting Corp's statistics suggest that 50 percent of people repeat a visit to the store every 33 days.

Churn happens mainly because of three reasons: change in customers' circumstances, the state of competition and change to product quality or features. Company can only control the latter two factors. Competitor can outsmart other brands by offering good with lower prices or better quality, or even both simultaneously. Strategic planners keep an eye on marketing mix of its competitors to avoid any customer diversion. Another potential reason for customer going to somewhere else can be because of deprivation in the product or service. Even lack of communication and poor information about one's customer can create cleavage between the buyer and the seller. Impacts of churning on the business growth are detrimental. It not only takes away market revenue of the brand but also affects its market value. For instance, if people stop buying on Tesco, it will generate a sea change for the company's reputation.

Focusing on the importance of the problem for Food Corp, this data driven study will help the company in preempting churning of customers. Given the time series data of active customers, predictive machine learning models have been applied on the temporal data set. It starts with creating temporal features for each customer given its unique receipts id. Monetary value of each visit with quantity inform about customer's basic purchasing behavior. Frequency of active customers is calculated using unique receipt id. An output window of 28 days is created to check output churn. A tumbling window of 7 days is observed to train the model for desired weekly predictions. This study is conducted on eight weeks period prior reference date. Categorical nature of the problem warrants classification model. Random forest and support vector machine are chosen for their simplicity and ability to take care of large features. To get most of these models, hyper parameter tuning is done on both models. Models are trained after finding optimal values of their parameters. This process with cross validation technique authenticates the accurate predictions. After carefully considering accuracy, precision, recall and F1 scores of the model, Support Vector Classifier (SVC) seem to be the best fit for the actual prediction of the churner in the next 28 days.

Based on churning prediction, insights draw factors appear to be important for both churning and loyalty. The model predicts more churners than non-churners. This will provide an opportunity for the company to introspect in advance. Those on the churning track are found to have less frequency in the past with lower monetary purchase. The pen-profiling of both classes offer insights to take preventive action. Food corp is advised to focus on customer persona in order to understand and offer plans specific to the customer. Lack of communication is found to be an issue for those who have bought in less quantities. The study also suggests potential incentives to tackle competitor's strategies.

The report focuses on short term planning to avoid churning. Even though it is good for predicting weekly churn, there is need to broaden the scope of this study for long term prediction with added complexity. The next study should incorporate anticipated competitors moves to predict complex but better results. Automating the churn reduction will help company in keeping its competitor at bay in the long run. Survival analysis and ensemble models seem to have potential for future modeling.

1. Current levels of churn

Consulting Corp report for Food Corp discusses facts about churn in general, and for Food Corp in particular. One can infer following information form analyzing figure one, two and three.

I. Before 87 days, more than 75 customers return to buy again. Those who don't are highly likely to never return as the curve flattens after this point. Keeping in mind world data for churning, this study has chosen 28 days to be the churning time. According to figure one, almost 48 percent of active customer would not, on average, to be considered to have churned. Thus, it is a good size of active customers to predict their weekly behavior.

II. As shown in figure three, Food Corp should expect to target at least 20.00 percent of active customers from this study as potential churners. This will be a good benchmark to evaluate our predictive classification models.

Choosing almost half of active customers for this study will not just be insightful but also cost efficient. It excludes non-necessarily large data of customers with definition of more than 28 days of churning. Besides costly marketing, it will save time by correctly predicting potential churners in the desired weeks.

## Part II. Technical Part

### 1. Moving forward with Raw Data

Among all the tables, two are most important for customer churn. Receipt and receipt_lines are primarily two tables used to create a merged table. Initially, a sum_table is created using receipt_id to identify customers with unique receipts. Sum of each receipt for quantities and values is calculated. Afterwards, sum_table is merged with receipts using receipt_id to make a final table, Merged_data, for churn prediction. While checking null values and any other discrepancy, there was no need for data cleaning.

### 2. Feature Selection from Temporal Data Set

Supervised learning, which needs labeled data, is used in this study for churn prediction. Therefore, features related to customer purchasing behavior are selected over a period of time. The merged table has temporal data set with features such as frequency, monetary, quantity and recency of each customer. Collectively, they unearth insightful information about the purchasing history of the customers with Food Corp. Using SQL magic and Merged table, a set of input and output features is created using output and tumbling windows. The former is used to save output churn after the reference date and last date of record. Whereas the latter has 8 periods of 7 day each to predict future churn of studied customers.

| | customer_id | f1_v | f1_q | f1_f | w1_shop | f2_v | f2_q | f2_f | w2_shop | f3_v | ... | f6_f | w6_shop | f7_v | f7_q | f7_f | w7_shop | f8_v | f8_q | f8_f | w8_shop |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | ... | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 |
| 1 | 2 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | ... | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 |
| 2 | 9 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | ... | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 |
| 3 | 14 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | ... | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 |
| 4 | 25 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | ... | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 |
| 5 | 28 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | ... | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 |
| 6 | 34 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | ... | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 |
| 7 | 37 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | ... | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 |
| 8 | 38 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | ... | 0 | 1 | 0.00 | 0 | 0 | 1 | 0.00 | 0 | 0 | 1 |
| 9 | 39 | 214.64 | 224 | 1 | 0 | 411.28 | 472 | 3 | 0 | 224.96 | ... | 2 | 0 | 234.24 | 336 | 2 | 0 | 164.16 | 184 | 2 | 0 |

Figure 1

The above figure shows input features for each customer. Four features per week for 8 weeks included to predict customer's churn. It shows monetary value(v), frequency(f) and quantity(q) of a customer's shopping in a particular week. Besides, shop features for each monitored period tells whether the customer came for shopping or not. As these features are changing every week, the model will get precise understanding of particular time. Tumbling window size of 7 days is used to get these features. However, output window size is kept at 28 to see whether the customer will churn in accordance with chosen definition. The reasons we have chosen eight periods is that adding another period is not improving accuracy.

The input and output features are divided into training, validating, and testing to combat overfitting.  The reason we have selected valid set is that it will not expose out test set to the trained model. These sets are chosen keeping in mind the reference date: Max(date) less output Windom size. Data focuses only on active customers by considering only those customers who have at least visited once in the output Windom time period. This selection will only target important customers and will avoid redundancy.

### 3. Model Selection

The whole purpose of this study is to find active customers who are the risk of churning. This makes a classification task with either customer will continue buying the Food Corp products or they will not return. Therefore, classification machine learning models will be used to predict target class, i.e., churners. Random Forest and Support Vector Machines are used in this study to classify shoppers.

Random forest is one of the best machine learning algorithms based on decision trees. It is chosen because it is less affected by outliers and can handle collinearity within features for accurate results. Besides, it helps in better feature selection by ignoring redundant features while splitting on important ones. It also accurately tells which input feature has higher or lower affect on outcome. Keeping in mind the business case of churn prediction, random forest will accurately tell which features have higher magnitude when it comes to churning. Moreover, it does not need standardization to perform better and can be easily deployed without any pre-scaling. Random forest can be accurate with high dimensional data as well and take care of variance.

Support Vector Classifier (SVC) is a Support Vector Machine classifier. This supervised method uses C optimization parameter to improve the boundary in hyperplane. SVC workss well with classification because of its ability to create decision boundary that extends the distance from closest data point of all the classes. The decision boundary created by SVM is also called Maximum margin plane. Even though it can take care of both linear and non-linear data, RBF kernel used in this study is for non-linear planes. It can take care of outliers and can separate different groups with more efficiency. Different gamma values are tried to take care of overfitting. Another reason for selecting SVC is that it handles complexity on its own and can handle thousands of features simultaneously. Having 12 features for our problem make it one of the best choices. However, it takes time on large data sets with large features.

### 4. Model Preprocessing

#### I. Standardization

Since there was no need for scaling data for Random Forest, it was not preformed while fitting the model. However, normalization is applied before fitting support vector machine due to non-normal structures of features. It is noticeable that standardization for SVC is done in way that we have used mean and standard deviation of training data set to scale validation and test data sets.

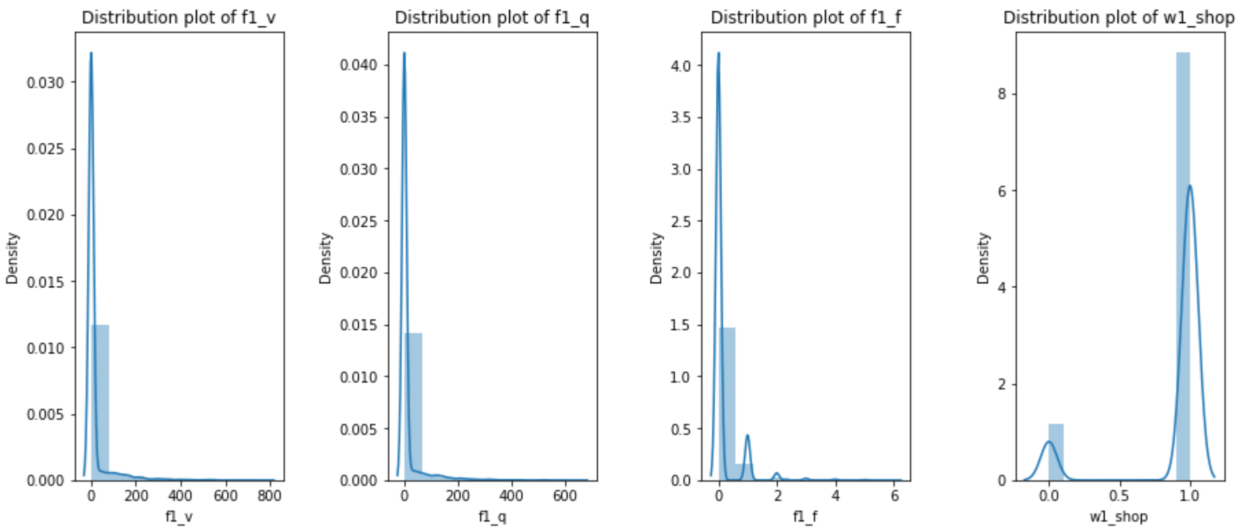


Figure 2

## II. Hyper Parameter Tuning/Optimization

Every machine learning model has parameters which need to be learned form data. However, certain parameters cannot be learned from regular training process. Improved parameters are called hyperparameters. For instance, it could be n_estimators and max_features for random forest. N_estimators show number of trees in the random forest. Optimization is a process of trail and error.

There are a few techniques for parameter tuning. Extensive grid search procedure is used before applying training data on our models. In this process, every combination of a range of values is tried to find the most optimal one. However, this takes time to find best parameters.

Figure 3 shows the results of Random Forest hyper parametrization for Random Forest. The best accuracy is observed with N_estimators(number of trees in the random forest) to be 100 with 12 max_features.

Similarly, optimization is done for SVM model. A range for gamma and C were along with kernel rbf (non-linear). These are important parameters in SVM model as they govern how lines are to be drawn in planes. Also, these depend on each other and were optimized at the same time. The optimum result for gamma and C came to be 0.01 and 100, respectively.

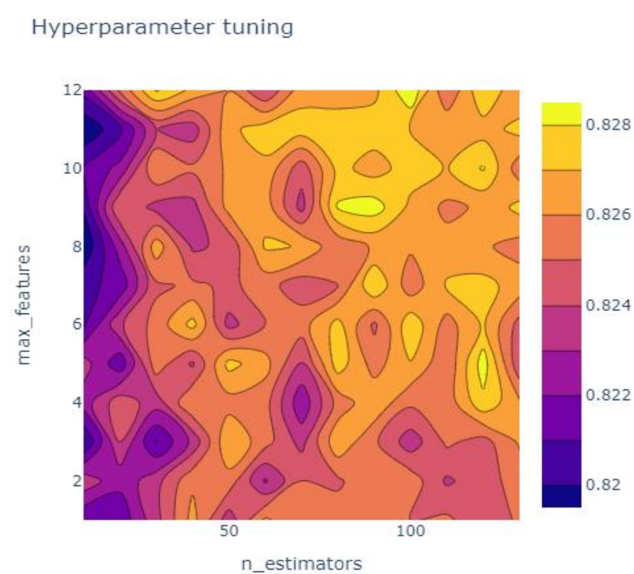


Figure 3

III. Cross Validation

Cross validation is a way we use our training data in order to get good estimates of how well our model will perform on data it has not seen before, i.e., test data.
Best estimator is using the test data, but its a good way to check before using test data. We use K-fold cross validation when we have un-balanced data set. With cv parameter as 5, cross validation confirms the higher accuracy of the model.

### IV. Feature Importance Using Univariate Feature Ranking

Using sklearn GenericUnivariateSelect method, these feature in figure 4 were ranked in accordance with their importance for predicting output. Not surprisingly, frequency top the list in predicting churn. Moreover, value seems to be important as well. Frequency of second last and fifth week form last purchase seem to be important for churn.

| Rank | Filter | |
|---|---|---|
| 0 | f5_f | : 0.0771 |
| 1 | f2_f | : 0.0733 |
| 2 | f8_v | : 0.0730 |
| 3 | f6_f | : 0.0721 |
| 4 | f4_v | : 0.0712 |
| 5 | f6_v | : 0.0709 |
| 6 | f2_q | : 0.0707 |
| 7 | f5_v | : 0.0699 |
| 8 | w4_shop | : 0.0699 |
| 9 | w2_shop | : 0.0689 |

Figure 4

## 5. Evaluation

Key performance indicators (KPIs) for model evaluation will be Accuracy, precision, recall and F1 scores. Accuracy is one of the most important indicators for correct prediction of both classes, i.e., churners and non-churners. Precision determines the record of accurate positive predictions. Recall, also called sensitivity, calculates the coverage of actual positive labels. F1 is the mean of recall and precision and tells overall story of the class predication.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.68 | 0.55 | 0.61 | 1302 |
| 1 | 0.86 | 0.91 | 0.88 | 3858 |
| accuracy | | | 0.82 | 5160 |
| macro avg | 0.77 | 0.73 | 0.75 | 5160 |
| weighted avg | 0.81 | 0.82 | 0.81 | 5160 |

Figure 5: Random Forest

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.77 | 0.46 | 0.58 | 1302 |
| 1 | 0.84 | 0.95 | 0.89 | 3858 |
| accuracy | | | 0.83 | 5160 |
| macro avg | 0.80 | 0.71 | 0.73 | 5160 |
| weighted avg | 0.82 | 0.83 | 0.81 | 5160 |

Figure 6: Support Vector Machine

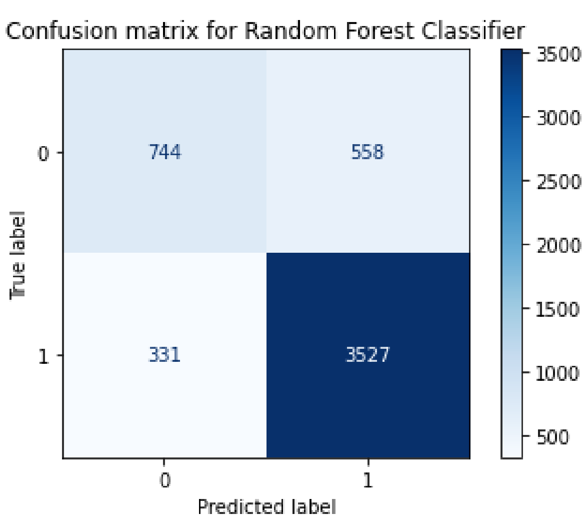


Figure 7

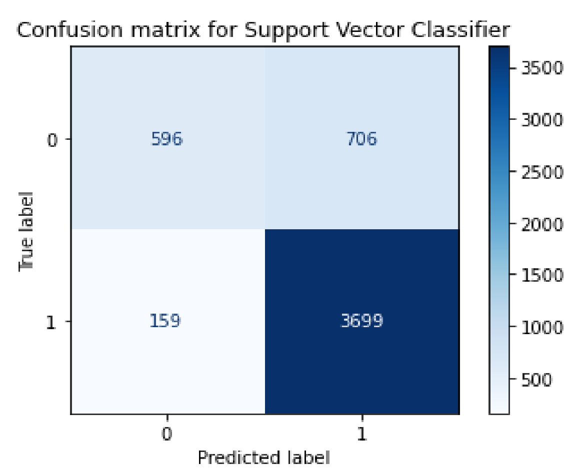


Figure 8

Accuracy shows overall true prediction of both classes by the model and is crucial for deciding which model to choose. However, if both of the classes are not equally important, then choosing important class in wrong or right way could be costly. For instance, precision is more important when the cost of false positive is high. For food Corp, cost of false positive is lesser than the cost of false negative. False positive for food Corp is someone who has been identified as churner while he is actually non-churner. It might cost to advertise to someone not churning, but the cost of losing a loyal customer by not advertising is substantial. Recall is important matrix as well as it tells whether the model has avoided making wrong classification. Having high precision and recall for the model is desirable. Balance of both precision and recall is what F1 score focuses on. When a model accurately predicts both classes, it will have higher F1 score. Comparing holistically, both are great models with robust accuracy.

**Summary**

Despite high performance by both models in all KPIs, we will choose Support vector classifier for two reasons. Firstly, it has higher accuracy, even by small margin of one percent. Second and more importantly, it has lower type II error, i.e., lesser false negatives as they are costly for Food Corp. Classifying more churners as non-churners(331 vs 159) is important reason for choosing support vector classification for further prediction.

## Part III: Insight Report

### I. Churn Predictions

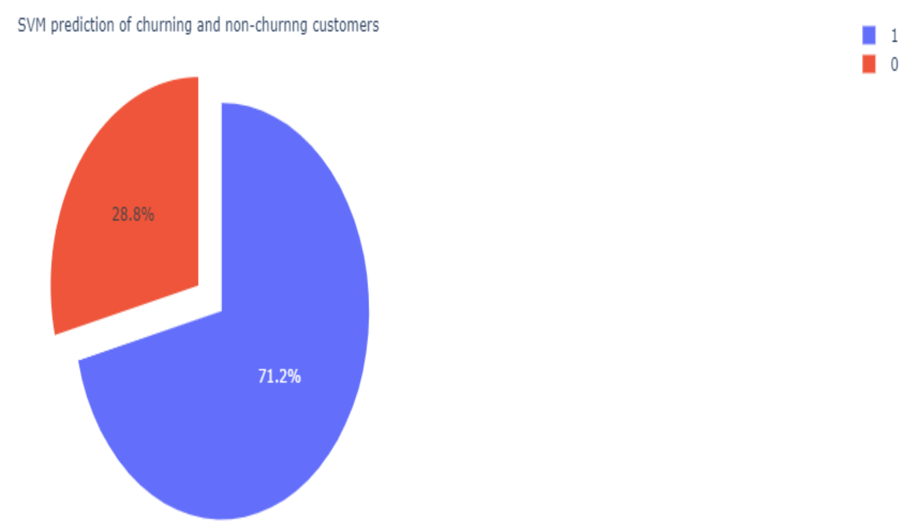


Figure 9

As support vector machine has finalized to be the best model for Food Corp data, it is applied to the future window of 28 days to predict churning and non-churning customers. The model predicts the above results. It predicts 3673(71.2%) as churning and 1487(28.8%) as non-churning customers in the next four weeks. These numbers should be a cause of concern for the company. Portraits of both classes are discussed in more datils.

II. Pen-Portraits

```
pen_pro_churn.describe()
```

| | customer_id | tot_val | tot_qty | frequency | Churn_predictions |
|---|---|---|---|---|---|
| count | 1120.000000 | 1.120000e+03 | 1120.000000 | 1120.000000 | 3673.0 |
| mean | 8394.469643 | 3.594873e+04 | 168.556250 | 13.539286 | 1.0 |
| std | 4715.733097 | 1.195222e+06 | 310.736144 | 25.964406 | 0.0 |
| min | 9.000000 | 1.150000e+00 | 1.000000 | 1.000000 | 1.0 |
| 25% | 4304.500000 | 1.716250e+01 | 14.000000 | 1.000000 | 1.0 |
| 50% | 8597.000000 | 5.757500e+01 | 40.000000 | 3.000000 | 1.0 |
| 75% | 12514.500000 | 2.475000e+02 | 171.250000 | 14.000000 | 1.0 |
| max | 16319.000000 | 4.000000e+07 | 2591.000000 | 298.000000 | 1.0 |

```
[244] pen_pro_nonchu.describe()
```

| | customer_id | tot_val | tot_qty | frequency | Churn_predictions |
|---|---|---|---|---|---|
| count | 472.000000 | 472.000000 | 472.000000 | 472.000000 | 1487.0 |
| mean | 8244.334746 | 234.371250 | 176.330508 | 13.580508 | 0.0 |
| std | 4569.601578 | 468.463993 | 391.457038 | 30.394328 | 0.0 |
| min | 0.000000 | 1.990000 | 1.000000 | 1.000000 | 0.0 |
| 25% | 4378.250000 | 15.687500 | 11.750000 | 1.000000 | 0.0 |
| 50% | 8172.500000 | 50.445000 | 38.000000 | 3.000000 | 0.0 |
| 75% | 12013.000000 | 238.552500 | 165.000000 | 13.000000 | 0.0 |
| max | 16285.000000 | 4433.120000 | 4589.000000 | 455.000000 | 0.0 |

Figure 10

Above description shows the key stats of churners and non-churners. As revealed in feature importance, frequency is crucial for determining the future relation of customer with the company. Those customers who have churned have an average frequency of 13.53 in the observed period against the average frequency of 13.58 for the non-churners. It is noticeable that churner frequency within one standard deviation of mean is substantially low, i.e., 25 for churners against 30 of non-churners. As quantity was also identified as another important feature for predicting output churn, it is evident that those who are non-churners have on average bought 10 more items than the churners. Given Food Corp is a retail business, number of quantity matters for both the store and the customer. Maximum quantity bough is also almost double for non-churners. The gap in the value of standard deviation for quantity is massive as non-churners have 80 more items in their standard deviation from the churners. Moreover, spending difference also signifies whether the customer will be loyal with the company or not. Those having low spending signal their weak relationship with the company and they are more at risk of leaving the company eventually.

### III. Business Insight for Food Corp Marketing Department

Academicians are time and again emphasizing the need to focus on retention. "It is now widely accepted that firms should direct more effort into retaining existing customers than to attracting new ones. " (AliTamaddoni, et al., 2014) In the contemporary world when information is easily available, customers are an easy prey for competitors. Therefore, Food corp must act proactively by foreseeing the customer future behavior. Given that a large proportion of active customers are at the risk of churn, it should think out of the box to retain its current customer base.

Firstly, marketing team of Food Corp need to target those customers who have been its active customer but buy less in quantity. A study by Harvard Business review customer experience team has found," Customers who are overspending and are pitched a better deal are much likelier to defect than over spenders who aren't approached." (Experience, 2015) Therefore, the management should equally target those active customer who buy less in quantity as suggested by its difference in churners and non-churners.

Secondly, general plans for everyone are not as much effective as specific ones are. Those who might go to another store should be offered pans that will serve their need. For instance, data from loyalty card can be used to analyze their consumption patten. They can be offered complementary vouchers or reward points for other brands as well if they spend certain amount in a time period. Therefore, customer-tailored strategy should be adopted to not just retain customers but involve their engagement with the brand.

Thirdly, saving deals have always been win-win for both businesses and customer. Food corp should compete with similar brand using the price game. Companies which save customers' money will incentivize its customer to stick with it. Therefore, value for money offers will help in increasing lifetime value of the customer.

Overall spending and loyalty are positively correlated. Marketing department should give incentives for higher spending. It will not come without clear message from the company. Offering exclusive deals with targeted communication will enhance Food Corp's value in the eyes of the customers.

Another tool marketing department can use to attract customers is clear communication. If a customer is not connected for a long period, he is not going to come back. Those old customers who have left should also be targeted with return incentives. People who leave Food Corp will go to another brand for grocery as it is necessity. Food corp should introspect its own policy. It should compare its price and promotion marketing mix strategies with that of its close competitor. It should make itself leader in the industry by adopting best practices to attract loyal customers.

### Conclusion

In sum, the study on Food Corp's data is revealing. The managers should analyze why their customers are churning in large numbers. Looking at their own data in the past, they will see they have neglected people who buy less. Moreover, the pattern suggests its competitor is either providing better goods at cheaper price, or the company is unable to communicate its product and value to the customers. Future models should also incorporate customer demographic and behavioral data to predict more accurate results. Complex machine learning models will help marketing department in understanding its customer in a better way. Monetary and other incentives with better communication to customer will have mind changing effects on potential churners.

## References

AliTamaddoni, J. S., Stakhovycha & Michael, E., 2014. Managing B2B customer churn, retention and profitability. *Industrial Marketing Management,* 43(7), p. 1258.

Experience, C., 2015. The Wrong Way to Reduce Churn. *Harvard Business Review,* p. 1.